Exercise Sheet 1

Exercises 3-5 from Jensen Pedersen Thomsen, Exercises 6-9 from Vaisman Zimányi

Sample Course Questions

1. Why have DataWarehouses been introduced. Do they replace traditional DBMS?
2. Differences between OLAP and OLTP
3. Describe the (4) components of multidimensional model (facts, hierarchy...)
4. Who queries the data warehouse. Through which tools/techniques?
5. Explain differences between the two main types of relational schemas in datawarehouses

1 OLAP vs OLTP

Exercise 1

OLAP vs OLTP queries

Indicate for each of the following queries if it would be typically executed on a data warehouse or operational database.

1.
```
SELECT Flight.ID, Flight.origin, MIN(price)
FROM Flight F, Company C, Airport A
WHERE F.CID= C.ID
  AND C.name='AIR FRANCE'
  AND F.origin = A.ID
  AND A.city='Paris'
  AND F.destination='JFK'
  AND F.arrival<'16.30'
  AND F.date='26/12/2014'
GROUP BY F.origin
ORDER BY price;
```

2.
```
INSERT INTO Flight
VALUES ('CDG','JFK',500€,'KLM',
'15.00','26/12/2014',AF15794);
```

3.
```
DELETE FROM Flight WHERE Flight.ID='AF755'
```

4.
```
SELECT orig, dest, compname, month, MIN(price)
FROM Flight F, Company C, Category Ct, Date D
WHERE F.CID= C.ID
  AND F.Ct= Ct.ID
  AND F.date_key=D.id
  AND D.year='2014'
  AND Cat='business'
GROUP BY F.orig,F.dest,C.compname,D.month
ORDER BY F.orig,F.dest,D.month;
```

Exercise 2

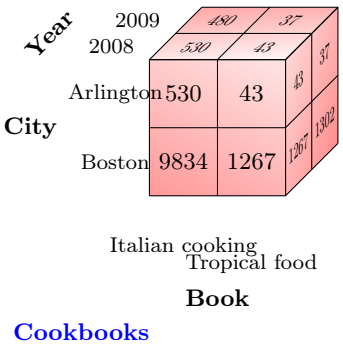
Cube, dimensions

1. Draw the dimension schemas for the cube Storesales from Figure 1, assuming we wish to investigate sales at month, quarter, year level for each city, country, by product subcategory (Beverage, Condiments, Seafood, Produces, TV, PC) and category (Electronics, Groceries).
2. How many granularities can be considered in that case?

Exercise 3

Cube

1. Illustrate how the cube Cookbooks on the right (measure: number of books sold) can be presented in a single spreadsheet nesting dimensions. (assume hidden cell has value 178)
2. Show how that cube would look after roll-up to $\top$ level in Location dimension.
3. Would you include the weather in the time dimension in this cube? And if the cube stores the sales of a single shop with an outdoor stand?



Cookbooks

Exercise 4

Facts taxonomy

For each of the following statements, explain what kind of a fact(event...) we are considering:

1. a fact exists for (d,p,c,s) if product p was sold to customer c from shop s on day d. The measure is total price.
2. a fact exists for (d,p,s) if product p was sold from shop s on day d. Measure is total price.
3. a fact exists for (d,p,s) for each product p, each shop s, and each day d. Measure is the total inventory of p in s for d.
4. a fact exists for (d,p,s) for each product p, each shop s, and each day d. Measure is the total amount of p sold in s in the year until day d.
5. a fact exists for (c,e,d) when a call from customer c was answered by employee e on day d. No measure.
6. a fact exists for (c,e,d) when a call from customer c was answered by employee e on day d. Measure is call's duration in seconds.

Exercise 5

Measure taxonomy

For each of the following statements, explain what kind of a measure(additive...) we are considering:

1. a fact exists for (d,p,c,s) if product p was sold to customer c from shop s on day d. The measure is total price.
2. a fact exists for (d,p,s) if product p was sold from shop s on day d. Measure is total price.
3. a fact exists for (d,p,s) for each product p, each shop s, and each day d. Measure is the total inventory of p in s for d.
4. a fact exists for (d,p,s) for each product p, each shop s, and each day d. Measure is the total amount of p sold in s in the year until day d.
5. a fact exists for (b,c,d) if currency *c* was exchanged in bank branch *b* on day *d*. Measure is the total amount exchaned in US $.
6. a fact exists for (b,c,d) if currency c was exchanged in branch b on day d. Measure is average exchange rate between c and US $ for all transactions involving c on day d.
7. a fact exists for each (b,c,d) where c is a currenct, b a bank branch, d a date. Measure is the total amount of c kept in b on day d.

2 OLAP operations

Another OLAP operation: UNION

- merge 2 cubes with the same schema (but different instances)
- can merge cubes with distinct granularity on one dimension
- syntax: UNION(Cube1,Cube2). Ex: AllSales←UNION(StoreSales,NordicSales) in Figure 1.

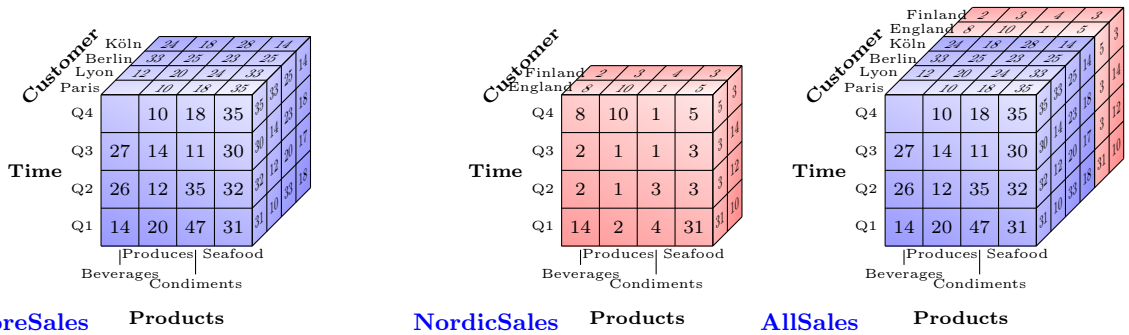


Figure 1: UNION of two cubes

Exercise 6

Multidimensional operations: basics. A data warehouse of a telephone provider consists of five dimensions,

namely, caller customer, callee customer, time, call type, and call program, and three measures, namely, number of calls, duration, and amount. Define the OLAP operations to be performed in order to answer the following queries. Propose dimension hierarchies when needed.

1. Total amount collected by each call program in 2012.
2. Total duration of calls made by customers from Brussels in 2012.
3. Total number of weekend calls made by customers from Brussels to customers in Antwerp in 2012.
4. Total amount collected from customers in Brussels who are enrolled in the corporate program in 2012.

Exercise 7

Multidimensional operations: basics. A data warehouse of a train company contains information about train segments. It consists of six dimensions, namely, departure station, arrival station, trip, train, arrival time, and departure time, and three measures, namely, number of passengers, duration, and number of kilometers. Define the OLAP operations to be performed in order to answer the following queries. Propose dimension hierarchies when needed.

1. Total number of kilometers made by Alstom trains during 2012 departing from French or Belgian stations.
2. Total duration of international trips during 2012, that is, trips departing from a station located in a country and arriving at a station located in another country.
3. Total number of trips that departed from or arrived at Paris during July 2012.
4. Average duration of train segments in Belgium in 2012.
5. For each trip, average number of passengers per segment, that is, take all the segments of each trip, and average the number of passengers.

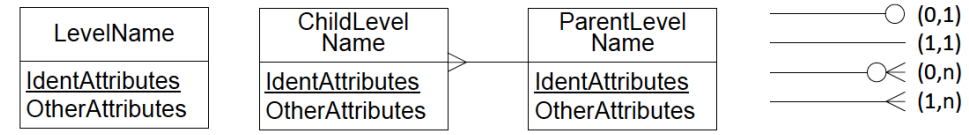
Exercise 8

Conceptual model design. Design a Multidim schema for the telephone provider application of Exercise 6

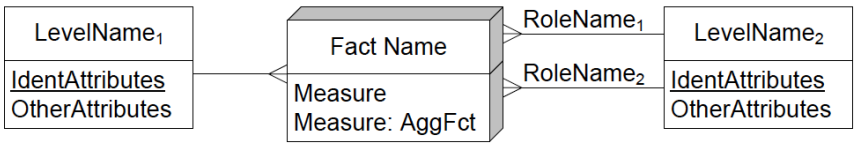
Exercise 9

Harder OLAP queries. Write using the OLAP operations the following queries on the Foodmart cube from Figure 2. Remark: we will adopt the following syntax to keep an attribute (ex: store manager) in the "group by" (this syntax amounts to viewing the attribute as a standalone dimension): `ROLLUP( ..., [Store manager]`→`[store manager], ...)`. We will also use the syntax `Measures.*` to describe the set of all measures.

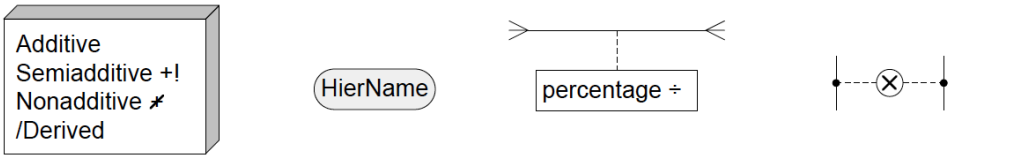
1. Sales average in 1997 by store state and store type.
2. All measures, including the derived ones, for stores in the state of California summarized at the state and the city levels.
3. Unit sales by promotion, year, and quarter.
4. Unit sales by city and percentage of the unit sales of the city with respect to its state.
5. Unit sales by city and percentage of the unit sales of the city with respect to its country.
6. Sales profit by month and percentage profit growth with respect to the previous month.



(a) Level (b) Hierarchy (c) Cardinalities

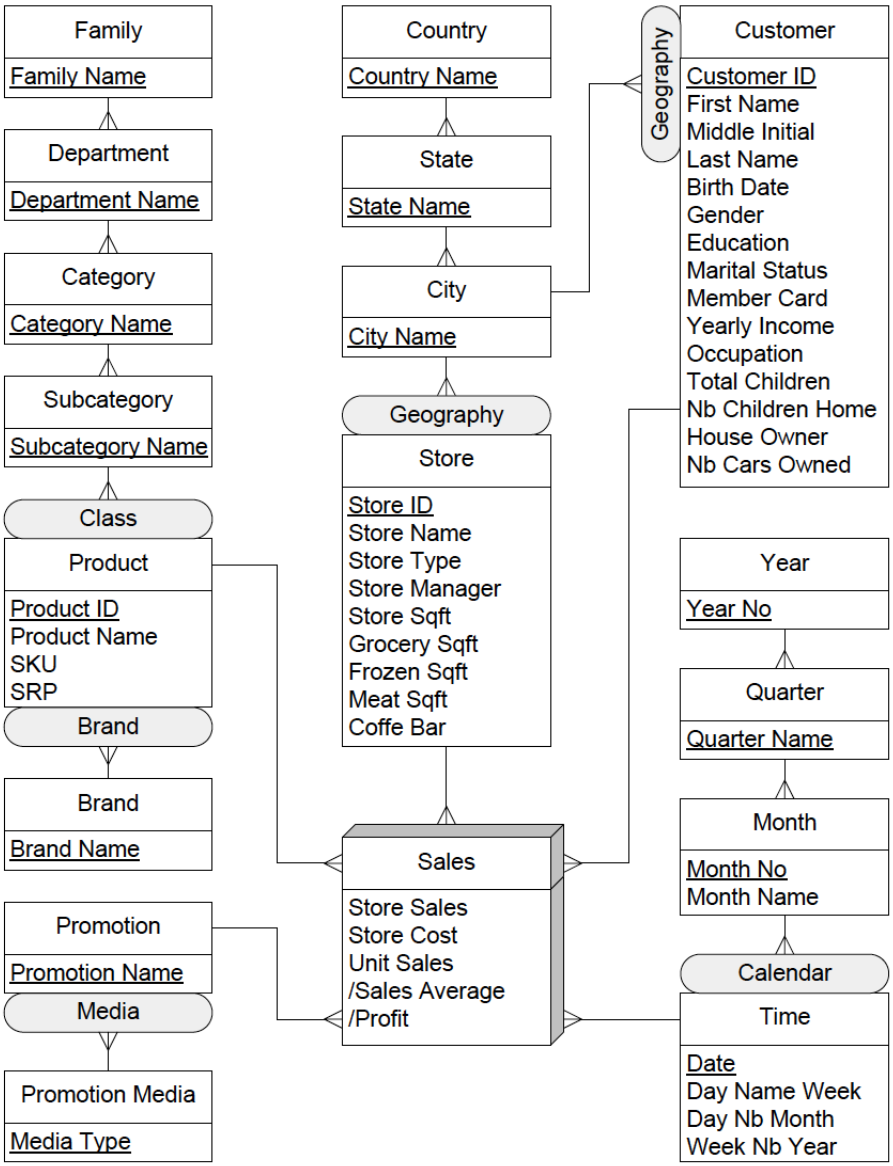


(d) Fact with measures and associated levels



(e) Types of measures (f) Hierarchy name (g) Distributing attribute (h) Exclusive relationships

Notations of the MultiDim model



MultiDim schema for Foodmart cube

Figure 2: A conceptual model for DW